Maulana Azad National Urdu University

Ph.D. (Computer Science) Coursework I Semester Examination, February 2022

مشین لرننگ

PHCS105DST: Machine learning

Time : 3 hrs **Marks : 70**

ہدایات:

یہ پرچہ سوالات دو حصوں پر مشتمل ہے: حصہ اول اور حصہ دوم۔ ہر جواب کے لئے لفظوں کی تعداد اشارۃً ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں **10** لازمی سوالات ہیں اس میں طالب علم کو کوئی 08 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً سو **(100)** لفظوں پر مشتمل ہے ہر سوال کے لیے 05 نمبرات مختص ہیں۔ **(8 x 5 = 40 Marks)**

2. حصہ دوم میں 05 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 03 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً ڈھائی سو **(250)** لفظوں پر مشتمل ہے۔ ہر سوال کے لیے **10** نمبرات مختص ہیں۔ **(3 x 10 = 30 Marks)**

حصہ اول

1. Model کی Performance کا جائزہ لینے کے لئے Confusion Metrics میں استعمال ہونے والے TPR اور FPR کو مثال کے ساتھ وضاحت کیجئے۔

2. Information gain (IDC/C4.5) کا استعمال کرتے ہوئے Attribute selection measure کا formula لکھئے اور مثال کے ساتھ وضاحت کیجئے۔

3. Learing System میں cyclic process ہے ان میں شامل بنیادی اقدامات لکھیے Learning Models کے Testing میں استعمال ہونے والے method کے بارے میں لکھیے۔

4. Classification اور Prediction کے حوالے سے مختلف مسائل کے بارے میں لکھیے۔ Back propagation method سے Prediction کی وضاحت کیجیے۔

5. Intelligent Model کی Performance کو کیسے پتہ ناپتے ہے؟ Confusion Matrix کی وضاحت کیجیے۔

6. ANN، SVM کے کیا فوائد ہیں؟ ANN میں overfitting کے مسئلے پر SVM کیسے قابو پاتا ہے لکھیے۔

7. Data preprocessing کے دوران normalization کو کیوں apply کرتے ہیں۔ مناسب مثال کے ساتھ وضاحت کیجیے۔

8. MPL کے Architecture کا خاکہ بنایئے Bias, learning rate اور Activation function کی وضاحت کیجیے۔

9. Support Vector machine (SVM) میں Maximum marginal Hyperplane (MMH) سے کیا مراد ہے؟ ANN اور SVM کے کیا فائدے ہے؟

10. Random forest algorithm کے اقدامات لکھئے۔ خاکے کی مدد سے Classification کے لئے Voting function کی بھی وضاحت کیجیے۔

حصہ دوم

.11 مندرجہ ذیل dataset کا استعمال کرتے ہوئے Decision Tree Induction Algorithm کے اقدامات کی وضاحت کیجیے۔ اس Algorithm کے فوائد اور نقصانات کے بارے میں لکھیے۔

Customer's Income	Customer Age	Credit Card rating	Expected outcome of booking Air-Tickets
high	<=30	fair	no
high	<=30	excellent	no
high	31...40	fair	yes
medium	>40	fair	yes
low	>40	fair	yes
low	>40	excellent	no
low	31...40	excellent	yes
medium	<=30	fair	no
low	<=30	fair	yes
medium	>40	fair	yes
medium	<=30	excellent	yes
medium	31...40	excellent	yes
high	31...40	fair	yes
medium	>40	excellent	no

.12 مندرجہ ذیل Dataset، Naive-Bayes Classification کا استعمال کرتے ہوئے ایک unseen sample، X=< NO, Single, 250000, Yes> کے output کو generate کیجیے۔

T_ID	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

.13 Similarity اور Dissimilarity سے کیا مراد ہے؟ اپنی پسند کی مناسب مثال کے ساتھ Clusters کے Partitional قسم کی وضاحت کیجیے۔ اور K-means clustering کے اقدامات لکھیے۔

.14 MPL Back Propogation کے Architecture کے خاکے کو بنایئے ANN میں استعمال ہونے والے مختلف activation function کے بارے میں لکھیے Gradient Descent Algorithm کی وضاحت کیجیے اور کوئی بھی تین Error Metrics کا حساب کیجیے۔

15. مندرجہ ذیل Dataset کے لئے Sales کا Target outcome اور دوسرے Input Variables کے درمیان Relationship کو قائم کرنے کے لئے Regression یا Classification Method کو Apply کرتے ہیں۔ یہ بتانے کے لئے کہ اقدامات لکھئے۔ (Steps only no Script) جو یہ ظاہر (determine) کرے کہ ان میڈیا (Media)، (TV & Radio or Newspaper) میں سے کون سا میڈیا (Media) کمپنی کو زیادہ سے زیادہ منافع دے گا۔

ID	Social Media	TV & Radio	Newspaper	Sales in m$
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9
6	8.7	48.9	75	7.2
7	57.5	32.8	23.5	11.8
8	120.2	19.6	11.6	13.2
9	8.6	2.1	1	4.8
10	199.8	2.6	21.2	10.6

☆☆☆